DOCUMENT RESUME

ED 343 928 TM 018 076

AUTHOR Klein, Thomas W.
TITLE A Proposal for Nevada's Participation in the National Assessment of Educational Progress Trial State Assessment.
PUB DATE 15 Oct 90
NOTE 13p.
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Comparative Analysis; *Cost Effectiveness; Data Collection; *Educational Assessment; Elementary Secondary Education; Longitudinal Studies; Mathematics Tests; National Surveys; Pilot Projects; *Program Proposals; Research Methodology; *State Programs; Testing Programs
IDENTIFIERS *Nevada; *Trial State Assessment (NAEP)

ABSTRACT

The original National Assessment of Educational Progress (NAEP) has been criticized for its limited ability to provide information that is useful at a variety of levels. As a result, Congress authorized a voluntary NAEP Trial State Assessment Program for the assessment of mathematics at grade 8 in 1990, reading at grade 4 in 1992, and mathematics at grades 4 and 8 in 1992. These assessments will permit state-by-state comparisons. Considerations relevant to the participation of Nevada in the trial assessment are reviewed, and a summary of expected costs and benefits related to participation is presented. The many expected benefits are significant in their promise for improving both statewide assessment and the longitudinal analysis of educational progress within the state. The estimated direct cost would be less than 12% of the current budget for state-mandated proficiency testing. It is recommended that Nevada commit itself to the 3-year trial participation. The expected costs and benefits are summarized in one table. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED343928

A Proposal for Nevada's Participation in the National Assessment of Educational Progress Trial State Assessment

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Thomas W. Klein, Ph.D.

October 15, 1990

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
K. CROWE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Introduction

The 1867 legislation which established the United States Department of Education provided for the federal monitoring of academic achievement and the reporting of results to the public. More than a century passed before an ongoing, uniform national system of academic assessment was established. That program is the National Assessment of Educational Progress (NAEP), which collected its first nationwide data in science and writing during the 1968-69 academic year. Since 1969, NAEP has assessed a wide variety of subjects including reading, mathematics, art, music, geography, social studies, citizenship, and career and occupational development in 9, 13, and 17-year-old students. NAEP has also carried out special assessments on limited samples, at specific ages and grades, in other subjects.

Among the most notable changes in NAEP since its inception, has been the movement away from age as a reference for its data. As originally conceived, NAEP would concentrate its efforts on students who were 9, 13, and 17 years of age, ignoring the grade level in which students were enrolled. This plan was partially abandoned when some later assessments were carried out for students enrolled in grades 3, 7, and 11. At present, most NAEP assessments focus on grades 4, 8 and 12, incorporating limited samples based on age in order to provide the necessary link among data collected at different times.

Federal legislation passed in 1988 provides for the bi-annual assessment of reading and mathematics, the assessment of science and writing every four years, and an assessment of history/geography at least every six years. It also provides for a bipartisan National Assessment Governing Board, a policy group with responsibility for the selection of subjects to be assessed within the framework established by legislation. This board has expressed its intent to expand both NAEP's scope and influence. Thus, the range of subjects assessed by NAEP can be expected to exceed the minimums specified by the legislation.

The original National Assessment of Educational Progress has been criticized for its limited ability to provide information that is useful at a variety of levels. Although the responsibility for public education is clearly vested in the states, NAEP data are reported only for the nation as a whole and for four large geographic regions. As a result, Congress authorized a voluntary NAEP Trial State Assessment Program for the assessment of mathematics at grade 8 in 1990, reading at grade 4 in 1992, and

TM018076

BEST COPY AVAILABLE

mathematics at grades 4 and 8 in 1992. These assessments will permit state-by-state comparisons and pressure is being brought to bear to extend the administrative units for which summaries are provided to the school district and even building levels. The 1990 and 1992 assessments were intended to be trial assessments, with a decision on continuation and expansion of this component of NAEP dependent on evaluation of the results. However, the National Assessment Governing Board has already moved to support full state participation and full federal funding of state-by-state NAEP.

Characteristics of NAEP Assessment

NAEP - National and State-by-State

The national assessment is intentionally designed to maintain the anonymity of its participants, both individuals and institutions. Student names are purposely excluded from test documents and results are not reported for political divisions below the multi-state region. This procedure minimized the stakes associated with participation. That is, neither rewards nor sanctions are associated with the school's, the district's, or the state's performance on the test. Two positive consequences of this approach are, first, the expectation that the program will provide an objective assessment of academic achievement, since no personal gain can be expected to result either from efforts to "fake good" or efforts to "fake bad." Thus, the deliberate falsification of data should not be a problem. Whether the lack of consequences affects student motivation to excel could be questioned. The second positive consequence is the assurance against the inadvertent release of individual data that these procedures provide. However, these benefits do not accrue without cost. The major cost associated with these limitations on National NAEP is a severe restriction of the utility of the data. Finer analyses of student progress are precluded and the relationship of student achievement to a variety of variables can only be studied to the extent that the effects generalize across large geographical areas and a variety of curricula. The research value of the data from National NAEP, beyond the description of regional demographic trends, is severely limited.

The limitations on use of National NAEP data was a major stimulus for congressional authorization of a Trial State Assessment Program (state-by-state NAEP) in 1988. The state-by-state program duplicates, to a large extent, the national assessment process with regard to the nature of the data collected. However, it has been freed from some of the strictures imposed on the national endeavor. Efforts are underway to maximize the utility of the data, within the limitations imposed by the enabling and related legislation. If these efforts are successful, data which could even include individual student identification could be made available to states for use in

efforts such as the calibration of state mandated tests with NAEP, while preserving the anonymity of individual participants. Other alternatives which would permit calibration of local measures are also being considered. If this breakthrough is achieved, states would be freed from total dependence on publisher's test norms, in those instances where commercial tests are used as part of a state program, and locally developed instruments could be calibrated to provide a national reference for their results. State-by-state NAEP data might also prove useful in the evaluation of experimental curricula and other efforts made to improve the education of Nevada's citizens.

Sampling, BIB Spiraling and Score Reliability

National NAEP may be considered to be efficient in that it provides relatively stable estimates of academic achievement for large groups while minimizing the time invested by individual participants, both students and institutions. This is accomplished through the selection of only a representative sample of schools to participate in each National assessment and the selection of a random sample of students within each participating school. Thus, schools in only 35 to 40 states are involved in each National assessment of a subject area. The same sampling approach has been taken in the Trial State Assessment. The program currently draws a sample of at least 2,000 students from approximately 100 schools within the smallest administrative unit for which the program will report results. In a state-level assessment, this would represent between 15% and 20% of Nevada's students for each subject area at each grade level tested and the involvement of a much higher percentage of Nevada's public schools. Thus, sampling in a State-by-State NAEP assessment would have only limited effect in minimizing demands on schools and students in Nevada,

Examination booklets used in both National NAEP and the Trial State Assessment represent a variety of forms which differ in content. Through a process termed BIB Spiraling, each student answers only a sample of items from the NAEP item pool during the test period. This minimizes the amount of testing time required for each student in the sample, but the scores for individual students have only limited reliability. Even building level results are quite variable and much less reliable for most schools than indicators provided by the Nevada Proficiency Examination Program. When student-level results are aggregated across a number of schools, a representative sample of students would be expected to have responded to each of the items in the pool and stable estimates of the achievement of the larger group are derived. This aspect of NAEP provides yet another illustration of an inherent characteristic of any large-scale assessment program. i.e. The least accurate and most error prone of the achievement estimates derived from the effort will be those scores derived for individuals. The comprehensive assessment of academic achievement across a variety of subject areas at the student level is an expensive

exercise in terms of both time and resources. To date, NAEP has made no pretense of adequately assessing the academic achievement of individual students. That has not been its purpose. However, using procedures such as sampling and BIB spiraling, federal agencies have been able to obtain adequate information for purposes such as accountability without significantly disrupting the local educational program. The testing burden on schools and districts will increase if adequate state-level data are to be obtained.

Scales Used In Reporting NAEP Results

Scales used for reporting scores on NAEP tests are developed through the application of Item Response Theory. These procedures provide comparable estimates of achievement for individuals even though students respond to different items. Scores in each subject area assessed can range from 0 to 500. The scales are so constructed that very few students would be expected to score at the extremes, either below 150 or above 350. The scale is continuous across the grade levels assessed by NAEP, and it is expected that about half of middle-school students will score above 250 on the examinations and about half will score below 250. Five anchor points are identified for each scale, one at each 50-point interval from 150 and 350. Each anchor point has a criterion-referenced interpretation, providing an example of the skills that a student scoring at that level would be expected to possess. These proficiency levels provide the reference for the reporting of results in terms of the percentage of students that score at or above each level. The meaningfulness of this form of reporting is dependent on the appropriateness of conceptualizing the subject matter as hierarchical in nature, with proficiency in those skills lower on the scale required for performance at the higher levels. The more hierarchical the underlying structure of the subject matter, the more informative and accurate is this criterion-referenced interpretation. NAEP was not designed as a criterion-referenced or mastery test. However, its description of proficiency levels in terms of specific skills contributes much to the comprehensibility of results. NAEP data are also reported in terms of the mean proficiency score for simplicity in making comparisons.

The National Assessment Governing Board has recently moved to establish three criterion levels of performance at each grade level assessed. These they have termed Basic, Proficient and Advanced. The Basic level represents minimum skills which should be expected of all students at the grade level. Proficient represents performance that is at or above grade level expectations and Advanced is used to describe exceptionally able performance. Groups of experts have been convened to define the types of behavior that represent these levels of performance for each grade, for the subjects tested in the Trial State Assessment. This effort took the form of asking experts to judge which test items represented which of these achievement levels, much as the current anchor descriptors for the cross-grade scales were

developed. Due to broad dissatisfaction with both the process and results of the initial standard setting conference, a second meeting was recently convened to refine the initial product. This setting of within-grade standards has been criticized as an attempt by the National Assessment Governing Board to unduly expand the influence of NAEP and to preempt efforts by the nation's Governors to define both appropriate educational goals and standards for the nation and the manner in which progress toward those goals are assessed.

Timeliness of NAEP Reports

National NAEP has been criticized for the time that elapses between the assessment and publication of results, the typical delay being between twelve and eighteen months. Critics of this phase of the program often cite the six to eight week turn around provided by commercial test publishers. Although the criticism may have some validity, the comparison with commercial tests does not. Commercial tests are scored on a predetermined scale which is developed and fixed prior to publication of the assessment instrument. In contrast, NAEP scales must be adjusted with each new assessment effort to maintain their continuity with earlier tests. While commercial test scoring services merely scan answer documents and report scores for individuals and summaries for specified levels of aggregation, NAEP carries out extensive analyses of the data relative to a variety of background variables and educational practices. These efforts, beyond those provided by commercial test publishers, are all quite time consuming, even when the time required to produce the printed report is ignored. Few institutions even undertake such a thorough analysis of data obtained from proficiency testing; thus, appropriate comparisons for judging the reasonableness of the delay are difficult to find. When compared to the time required by academic studies of similar magnitude, the twelve-to-eighteen-month delay does not seem unreasonable. The National Assessment Governing Board seems particularly sensitive to this issue and is moving to determine if the timeliness of NAEP reporting can be improved.

The Future

Trends in national assessment may be difficult to predict due to the confusing and, in some cases, contradictory administrative structures that influence NAEP. President Bush has expressed educational goals that include the testing of all fourth, eighth, and twelfth grade students in the country. The grade levels included in those remarks suggest an expansion of NAEP. However, such an expansion would require significant changes in the purpose as well as the practice of national assessment as reflected in NAEP. We will have to wait to see, first, if he meant what he said, and second, whether he can gather the political support that would be required for such an effort.

Clearer evidence of probable direction comes from the National Assessment Governing Board. At its December 9, 1989, meeting , the board adopted "Positions on the Future of the National Assessment" which included support for:

1. Assessing at least three subjects each year

2. Full state participation and full federal funding

3. Removal of prohibitions against use of NAEP tests and reporting of results below the state level

4. Establishment of international samples to participate regularly in the assessment

5. Decreasing both the development time required for NAEP exams and the delay in reporting the results

6. Revision of NAEP exams to include the full range of knowledge and skills, decreasing the perceived emphasis on basic skills and including more advanced material

7. Increasing the samples tested to provide reliable information on additional subdivisions of the larger population

8. Clarification of the governing structure for NAEP

The governing board has been criticized for adopting these far-reaching proposals without seeking extensive consultation with the states and other interested groups. The ambitiousness with which the Board has moved to exert its influence on NAEP, particularly the Trial State Assessment, has also been seen by many as cause for concern. A number of efforts have been mounted both by private organizations and the United States Department of Education to question the wisdom of these proposals which would both markedly expand and redefine NAEP.

The original National assessment earned a well-deserved reputation as a quality program that yielded valuable data while limiting its intrusion on the established educational system. If the goals stated above are to be realized, NAEP must change in dramatic fashion. It would not be appropriate to assume that such an expanded program, which might share only its name with the original, would deserve the same enthusiastic support.

Some Considerations Relevant to Nevada's Participation in State-by-State NAEP

Expected Benefits

The National Assessment of Educational Progress is a quality program with limited goals. It could be argued that the program has been able to maintain that quality because its goals have been limited. The "low-stakes" nature of NAEP participation may have also contributed to the maintenance of quality while promoting the broad acceptance of NAEP's efforts, even among many of the most vociferous critics of standardized testing. At the present time there seems to be hardly an educator who cannot find something about NAEP to love. When state-by-state NAEP results are first reported for the 1990 Trial State Assessment, this near uniform acceptance can be expected to decline. Despite the various efforts to alter the nature and purpose of NAEP, it seems reasonable to expect that its quality will be maintained, at least in the intermediate term (3-5 years).

Given NAEP's excellent reputation among educators, the major impediment to participation in state-by-state NAEP would seem to be its cost relative to perceived benefits. If the only benefit were to be the opportunity of having Nevada ranked among the other participating states on the subjects assessed, it could be argued that its cost would be excessive. The prospect that the data will be fully shared with the states, for such purposes as calibrating locally administered tests, provides an entirely different dimension favoring participation. The calibration of proficiency examinations administered at grades 3, 6, and 9 would not only provide an additional national norm reference for the results but would promote the longitudinal analysis of proficiency test data despite necessary changes in test forms, editions, and/or even publishers. The ability to reliably track the progress of education within the state of Nevada would be greatly enhanced. The comparison of the performance of LEAs on NAEP and on the proficiency examinations would provide a recurring check on the validity of the proficiency examination results. NAEP would also be available to provide a national norm reference as well as a quality validating criterion for proficiency examinations developed by the State Department of Education.

These benefits could be realized while maintaining the assurances of anonymity in the publication of results. The association of the student's name with his/her data might be required at the level of data collection and analysis; however, there would be no requirement for the release of that information in any published reports and prohibitions against such release would be stringently enforced.

It can be expected that the major utility of even state-by-state NAEP data will be at the level of the large group. Thus, NAEP, or local assessment instruments calibrated to NAEP scales, would preserve their value in the evaluation of changes in curricula. However, attempts to force NAEP to yield student-level information, which might be useful in instruction, can be expected. A change of this magnitude could be expected to drastically alter the nature of NAEP for little conceivable gain. Such an expansion in the scope of NAEP would not only drastically increase its cost, but it could be expected that the result would suffer from the same shortcomings as nationally normed commercial tests. In their attempts to provide information that would be of some value across a variety of curricula and local conditions, most objective-referenced scales derived from commercial tests fall far short of the potential for diagnosis and prescription of instruction possible using measures specifically tailored for the curriculum in use. The comprehensive assessment of the academic skills of individuals, for the purposes of diagnosis/and prescription of instruction might be a task most prudently left to the local educational agency, where the assessment can be tailored to address those issues relevant to the local curriculum and local conditions.

In considering the issue of timeliness, the purpose of the assessment should not be ignored. Timeliness is generally a major consideration when the purpose of the assessment is to effect changes in the educational programs of individual students. If, however, the purpose of state-by-state NAEP is to report on the status of institutions, whether the report is at the national, state or district level, a delay of a year between data collection and reporting would rarely, if ever, be expected to prove critical for decision making processes which would utilize those data. In terms of the uses which states would have for the data, should the full sharing of those data materialize, the consequences of the delay could be ameliorated through release of the data to state agencies as soon as the work on scaling had been completed. States could then begin work on such projects as calibrating local measures with NAEP, while awaiting the federal report. Timeliness in the handling of the data for those purposes reserved to the states might then be satisfactory.

Providing an Effective Voice for Nevada in the National Debate

As indicated above, this is a time of intense activity aimed at redefining the National Assessment of Educational Progress to increase its utility for federal policy makers and provide information that would be valuable to the state. It seems reasonable to expect that this will be accomplished in much less dramatic fashion than that envisioned by the National Assessment Governing Board. Implementation of their proposals could well result in a prohibitive testing burden for a state the size of Nevada. For these reasons, it is imperative that Nevada provide for its input into the process that will define NAEP's future.

Participation in the Trial State Assessment would accomplish two purposes. First, it would insure Nevada's full participation in the NAEP Network, the group of state representatives that convenes several times each year at NAEP expense to review NAEP activities and consider plans for the future. The second purpose would be to provide the state with the experience necessary to reach an informed decision in regard to its future involvement in the State-by-State assessment should the Trial State Assessment be judged to be a success and funds be appropriated for its continuation and expansion to other grades and subject areas. The input and opinions of a participant in the Trial State Assessment, into the evaluation of the trial, could also be expected to carry more weight than those of an outside observer.

Assessment will drive instruction. Even a "low-stakes" assessment instrument can be expected to influence instruction by illustrating expectations for teachers and administrators. As the stakes increase, it may be anticipated that the pressures to meet those expectations will increase and the specificity of the skills that become the target for instruction will narrow. An assessment system, such as that envisioned by the Governing Board, one that spans a broad range of grade levels and extends beyond an emphasis on basic skills, could force curricula into a restrictive mold that would not be expected to serve the best interests of the diverse populations represented in Nevada. Whether determining the nature and ordering of instruction that takes place in our schools is an appropriate task for the federal government is a topic currently being debated in relationship to the proposed expansion of NAEP. It should be possible to effectively argue that the districts and states are in a much better position to make such decisions.

A second forum for representing the interests of the state in this area is the Assessment Subcommittee of the Education Information Advisory Committee of the Council of Chief State School Officers. This is the group most likely to influence NAEP policy as the debate on the Governing Board's proposals continues.

Regardless of one's position on the future of NAEP, it should be more effective to promote Nevada's interests in these decisions through participating both in the Trial State Assessment and the Education Information Advisory Committee rather than as a voice criticizing from without.

Expected Cost

It is estimated that the direct cost of Nevada's participation in the 1992 Trial State Assessment would be approximately $30,000. These funds are required to support the statewide coordination of the trial assessment; to fund the travel of the state coordinator and local test administrators to training sessions; to fund the travel expenses of the state coordinator to assist schools where needed; and to meet

necessary printing, mailing, photocopying and telephone expenses. Districts and schools that volunteer to participate in this program would be expected to contribute the time of their local administrators to this effort. The personnel costs to the agency and to LEA's that volunteer their participation should not be ignored. NAEP related activities can be expected to require approximately a quarter-time effort by the state coordinator. Each local administrator at the approximately 100 participating schools can expect to devote three to four days to NAEP related activities. In addition, principals and teachers of the subject areas assessed in the participating schools can be expected to be asked to complete background questionnaires, their responses to which will be related to student performance. Thus, the indirect costs can be expected to exceed the direct dollar amount estimated above.

The federal government's expenditures on behalf of each state's participation in the 1990 trial assessment has been estimated to be approximately $200,000 per state, several times the state/district/school expenditures.

The costs in subsequent years might not increase markedly, despite the proposed expansion of grade levels and subjects assessed, due to use of previously trained personnel.

Summary

A summary of expected costs and benefits related to Nevada's participation in the 1992 Trial State Assessment are presented in Table 1. The expected benefits are many and significant in their promise for improving both statewide assessment and the longitudinal analysis of educational progress within the state. The estimated direct cost, at the level of implementation of NAEP included in the 1992 trial assessment, would be less than 12% of the current budget for the direct costs of state-mandated proficiency testing. The indirect costs to both the agency and those LEA's that volunteer their participation are significant but necessary to insure the quality of the assessment. The expenditure would seem to be a timely investment.

Recommendation

State-by-state NAEP is a trial assessment program at this time. Thus, the strength of a recommendation for participation must be based on a variety of assumptions such as, the ability, under the law, of the National Assessment Governing Board to share NAEP data with the states and the timeliness with which those data might be made available. Under these conditions, it would seem to be a prudent course for Nevada to commit itself to a three-year trial participation in State-by-State NAEP. A three-year period would be required to fully assess the actual benefits of

participation but would limit the total resources committed to the project until those benefits could be fully evaluated.

The State Board and State Department of Education should also insure the availability of those funds required for travel of at least one Nevada representative to meetings of the Education Information Advisory Committee.

Table 1.

Summary of expected annual costs and benefits of Nevada's participation in the National Assessment of Educational Progress Trial State Assessment.

Expected Benefits

1. Participation in a quality assessment program, broadly accepted by the educational community, that will enhance the state's ability to track educational progress at three grade levels and is efficient in terms of its demands on student time.

2. Nevada will receive a statewide profile of the area assessed which will include the reporting of the relationship between achievement and a variety of school, curricular, and teacher variables.

3. The normative data will be current and provide a basis for comparison of Nevada's results with those of other participating states, territories, districts, and the nation on the measure that is expected to replace the less representative ACT and SAT scores as the anchor for the U.S. Department of Education's "wall chart".

4. The state will have access to a continuing quality measure to which state mandated proficiency examinations may be calibrated to provide continuity in assessment as editions and forms of locally adopted measures change.

5. NAEP data could be used to provide an ongoing validity check for proficiency examination results at the district and building levels if existing restrictions on the aggregation/reporting of NAEP data for these administrative levels are eased.

Expected Costs

1. Annual expenditure of approximately $30,000 in state funds to support state coordination, local administrator training, and other required activities.

2. One-quarter time effort by the state NAEP Coordinator.

3. Indirect costs of three to four days effort in NAEP related activities by each of the approximately 100 local NAEP administrators at those schools that volunteer to participate.

4. Time required of the principals and teachers of the subjects assessed, at the participating schools, to complete NAEP background questionnaires.